Wolfgang Kirk

Good Practice - 10: Dokumente erstellen, digitalisieren und verwalten

Good Practice - 10: Dokumente erstellen, digitalisieren und verwalten

- 1 Vorbemerkung
- 2 Markdown: Text einfach erstellen
- 3 Pandoc: Text einfach konvertieren
- 4 Typora: Text einfach konvertieren
- 5 Checkliste *Text erstellen und Konvertieren*
- 6 Digitalisierung
  - 6.1 Ausstattung
    - 6.1.1 Gerätetreiber und Software
    - 6.1.2 Alternative
  - 6.2 Einsatz
    - 6.2.1 Dokumente digitalisieren
    - 6.2.2 Inhalt extrahieren
      - 6.2.2.1 Copy and Paste
      - 6.2.2.2 Bordmittel
      - 6.2.2.3 Online
  - 7 Checkliste *Digitalisierung*
- 8 Literaturverwaltung
- 9 Checkliste *Literaturverwaltung*
- Impressum

1 Vorbemerkung

Wer Texte etc. erstellen oder alte Dokumente digitalisieren bzw. verwalten will, der kennt auch das Problem: Viele Lösungen, aber nicht immer abgestimmt auf die eigenen Bedürfnisse.

In der Praxis werden heute verschiedene Betriebssysteme genutzt. Von Android, Linux über MacOS bis hin zu Windows. Niemand kann da alle Möglichkeiten genau kennen.

Eine sinnvolle Lösung wäre, wenn die benutzten Hilfsmittel unter den gängigsten Betriebssystemen zur Verfügung stünden. Also Gerätetreiber und Software für Linux, MacOS und Windows.

Für Android stellen einige Gerätehersteller spezielle Texterstellungs, Druck- und / oder Scan-Apps zur Verfügung.

2 Markdown: Text einfach erstellen

Wer Texte in verschiedenen Umgebungen nutzen will, braucht nicht unbedingt eine Textverarbeitung. In diesem Fall ist die Auszeichnungssprache Markdown hilfreich.

Das oberste Ziel hinter Markdown ist einfach: Der Quelltext soll für **Menschen** lesbar sein. Texte sollen also keine Tags und kryptischen Formatierungen enthalten. Dieser Fokus auf Lesbarkeit hat den positiven Effekt, dass sich Markdown-Texte auch äusserst komfortabel verfassen lassen. Mehr Informationen sind erhältlich auf der Seite des Erfinders John Gruber unter **https://daringfireball.net/projects/markdown/** [2020-07.09].

Mit Markdown lassen sich beispielsweise alle grundlegenden Elemente eines Webtextes auszeichnen. Von Überschriften über Formatierungen wie *kursiv* oder **fett** bis hin zu Listen und Zitaten. Die Syntax ist denkbar einfach. Eine genaue Übersicht enthält die vorgenannte Internetseite von John Gruber.

Tipp

> Die Erstellung von Markdown-Texten ist auch mit jedem einfachen Texteditor möglich.

Markdown-Texte - erkennbar an der Abkürzung *.md im Dateinamen, werden üblich mit sogenannten Editoren erstellt. Es gibt eine Vielzahl derartiger Programme. Auf den komfortablen Editor mit Namen *Typora* wird hier hingewiesen, weil er für die Betriebssysteme Linux, MacOs und Windows verfügbar ist.

3 Pandoc: Text einfach konvertieren

Ein universeller Konverter ist *Pandoc*. Eine Liste der unterstützen Formate - darunter epub, html und pdf - ist aufgeführt unter **Pandoc.org**. Dieser Konverter ist ebenfalls für die Betriebssysteme Linux, MacOs und Windows verfügbar.

Das Programm wird über die Kommandozeile bedient. Mit ein wenig Einarbeitung in die Befehlssyntax lassen sich recht schnell gute Ergebnisse erzielen.

Eine Enführung bietet das Magazin Phlow unter **https://phlow.de/magazin/terminal/pandoc/** [2020-07-09]

4 Typora: Text einfach konvertieren

Der Markdown-Editor *Typora* bietet ebenfalls die Möglichkeit, den erstellten Text in andere Formate zu konvertieren. Derzeit werden folgende Formate unterstützt: pdf, html, docx, odt, rtf, epub, LaTeX, MediaWiki, rest, Textile, OPML, Bilder [2020-07-09].

5 Checkliste *Text erstellen und Konvertieren*

1. **Einfachheit**
    - Simpel: Text-Editor reicht grundsätzlich aus
    - Schnell: Texterfassung auch unterwegs, z.B. Smartphone, Tablet
2. **Kompatibilität**
    - Konvertierung: in viele gängige Formate
    - Cross-Platform-Software: auf vielen Endgeräten nutzbar
3. **Kosten**
    - Closed-Source-Software: Lizenz
    - Open-Source-Software: freie Nutzung

    Tipp

    > In vielen Fällen ist Open-Source-Software für die zu erledigende Aufgabe genauso gut geeignet!

4. **Zweck**
    - Aufwendiges Layout: Office Programm oder sonstige Spezialsoftware
    - Mehrfach-Nutzung: Einfache Auszeichnungssprache, z.B. Markdown.

    Tipp

    > Wissenschaftliche Texte schreiben - mit Markdown und Pandoc: siehe Beitrag von Justus Holzberger unter **https://vijual.de/2019/03/11/artikel-mit-markdown-und-pandoc-schreiben/**

6 Digitalisierung

6.1 Ausstattung

6.1.1 Gerätetreiber und Software

Beim Kauf von Druckern und Scannern ist darauf zu achten, dass die Hersteller für die o.a. Betriebssysteme spezielle Bearbeitungssoftware und die benötigten Treiber zur Verfügung stellen.

6.1.2 Alternative

Das Programm Master PDF Editor kann auch in der nicht-registrierten Version ohne viele Einschränkungen genutzt werden.

Wer bei den kostenlosen und / oder kostenpflichtigen Programmen zur Digitalisierung für seine Bedürfnisse nicht fündig wird, dem ist zu empfehlen, sich das kostenpflichtige Programm *VueScan Scanner Software* genauer anzusehen.

Beide Programme sind für Linux, MacOS und Windows erhältlich.

6.2 Einsatz

6.2.1 Dokumente digitalisieren

Üblicherweise werden Dokumente mit fast ausschließlich textlichem Inhalt als PDF-Dokumente, mit bildlichem Inhalt als PNG-Dateien elektronisch erfasst.

Darauf zu achten ist, das PDF-Dokumente als PDF-OCR-Text zu erfassen. Soweit möglich, sollte *Automatische Schiefe* als Korrektur eingestellt sein.

6.2.2 Inhalt extrahieren

6.2.2.1 Copy and Paste

Oftmals ist es einfacher, den Inhalt des Dokumentes zu kopieren und in einen Text-Editor oder eine

Textverarbeitung einzufügen.

6.2.2.2 Bordmittel

Soweit dies nicht zufriedenstellend gelingt gilt: An erster Stelle sollten die vom jeweiligen Betriebssystem bereits bereitgestellten Programme zur Konvertierung genutzt werden: LibreOffice, Evince, Gimp, ImageMagick, Scribus, etc.

a) Linux

Mit dem Tool *pdftotext* (Linux) lassen sich PDF-Dokumente in txt-Dateien konvertieren
Unter Ubuntu erfolgt die Installation in einem Terminal:

```
sudo apt-get install poppler-utils
```

Das Programm wird auf der Kommandozeilen-Ebene ausgeführt:

```
pdftotext -layout input.pdf output.txt
```

Optionen werden angezeigt:

```
pdftotext
```

Mit dem Dokumentenbetrachter *Okular* kann die PDF-Datei auch als txt-Datei abgespeichert werden. Die Installation erfolgt aus den Paketquellen:

```
sudo apt-get install okular
```

Menü **Datei/Exportieren als** öffnen und *Einfacher Text* ausführen.

b) Windows

Für Windows gibt es vergleichbare Programme:

PDF2TXT: **https://github.com/jamalmazrui/PDF2TXT** [2020-07-06]

Evince for Windows (allerdings in einer älteren Version)

https://www.chip.de/downloads/Evince-Document-Viewer-Portable_72939340.html [2020-07-06]

6.2.2.3 Online

Insbesondere bei besonderem Layout des Textes, z.B. mehrere Spalten, können Online-Konvertierer gute Ergebnisse erzielen.

Tipp:

> Zurückhaltung bei Dateien mit sensiblem Inhalt ist bei dieser Form geboten!

Text:

https://pdftotext.github.io/ [2020.07.06]

Hinweis:

Kein Upload erforderlich. Konvertierung geht sehr schnell. Aber: Besser den Text mit Copy and Paste in einen Text-Editor einfügen.

https://dokument.online-convert.com/de/umwandeln-in-txt [2020.07.06]

Bilder

https://bild.online-convert.com/de [2020.07.06]

7 Checkliste *Digitalisierung*

1. **Hardware**

   Software und Treiber sind für die Betriebssysteme Linux, MacOS und Windows erhältlich

2. **Software**

   1. Copy and Paste vor einer Konvertierung ausprobieren/nutzen
   2. *Bordmittel* des jeweiligen Betriebssystems nutzen, z.B. LibreOffice, Evince, Gimp, ImageMagick, Scribus, etc.
   3. Proprietäre Software nutzen: Master PDF Editor, VueScan Scanner Software, etc.

3. **Online**

   Die Online-Konvertierung sollte nur bei Dateien genutzt werden, die persönlich niemanden kompromittieren. Bei Dateien mit sensiblem Inhalt ist Zurückhaltung geboten!

8 Literaturverwaltung

Um den Überblick über selbst veröffentlichte Dokumente und die erworbene Literatur zu behalten, ist eine Verwaltung manchmal unerlässlich. Für die Fertigung von (wissenschaftlichen) Arbeiten wird in vielen Bereichen sowieso ein Literaturverzeichnis gefordert.

Als Grundlage bietet sich eine Literaturverwaltung im Format *BibTeX* an, weil diese Form auch in vielen (professionellen) Literaturverwaltungs-Programmen verarbeitet werden kann (Citavi, Zotero, etc.).

Beim Erstellen einer Bibliothek hilft das freie Programm *JabRef*, das für die Betriebssysteme Linux, MacOs und Windows verfügbar ist. Die erstellte Bibliothek ist eine einfache Textdatei mit der Endung *.bib, die auch mit jedem beliebigen Text-Editor bearbeitet werden kann.

Als Beispiel ist der Eintrag für dieses Werk im QR-Code unten angefügt.

9 Checkliste *Literaturverwaltung*

1. **Einfachheit**
    - Simpel: Text-Editor reicht grundsätzlich aus
    - Schnell: Texterfassung auch unterwegs, z.B. Tablet
2. **Kompatibilität**
    - Konvertierung: in viele gängige Formate
    - Cross-Platform-Software: auf vielen Endgeräten nutzbar
3. **Kosten**
    - Closed-Source-Software: Lizenz (z.B. Citavi)
    - Open-Source-Software: freie Nutzung (JabRef, Zotero)

    Tipp

    > In vielen Fällen ist Open-Source-Software für die zu erledigende Aufgabe genauso gut geeignet!

4. **Zweck**
    - Persönliche Literaturverwaltung
    - Mehrfach-Nutzung: z.B. als Literaturverzeichnis

    Tipp

    > Regeln für Zitate und bibtex-Einträge (ein Vorschlag zur Güte): siehe Beitrag der TU Dortmund unter **https://ls11-www.cs.tu-dortmund.de/de/teaching/regeln_bibtexeintraege**

Impressum

Verlag Wolfgang Kirk, Essen
ISSN 2627-8758
ISBN 978-3-96619-127-2, doi: 10.2441/9783966191272 (PDF)
ISNI 0000 0004 5907 4303
©2020 Wolfgang Kirk

Der Text ist als Band 57 Teil von Veröffentlichungen in der Reihe *Digitale Gesellschaft in Deutschland* (ISSN 2627-8758 elektronische Publikationen). Durch Auflösung der DOI auf der Seite https://dx.doi.org/ kann die elektronische Datei heruntergeladen werden.

Textsatz mit Typora in Markdown und mit Pandoc in das Zielformat konvertiert.

Stand: 2020-07-10

Haftungsausschluss

Der Autor haftet insbesondere nicht für den Inhalt der vorgestellten Internet-Seiten. Die Verantwortung für Inhalt und Funktion der Links liegt bei den jeweiligen Betreibern.

Für die Funktionalität der hier vorgestellten Programme und gemachten Angaben kann keine Garantie erfolgen. Auch wird eine Haftung für Schäden an der Hardware durch die Nutzung dieser Hinweise ausgeschlossen. Insgesamt handelt der Nutzer auf eigenes Risiko.

Lizenz

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz. Ausgenommen von dieser Lizenz sind alle Nicht-Text-Inhalte wie Fotos, Grafiken und Logos.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet über https://dnb.de abrufbar.

BibTeX